DOCUMENT RESUME

ED 230 104 HE 016 165

AUTHOR Cockburn, Barbara; Ross, Alec
TITLE Inside Assessment. Teaching in Higher Education Series: 7. Suggestions for the Consideration of Lecturers and Others Concerned with Teaching in Higher Education.
INSTITUTION Lancaster Univ. (England). School of Education.
REPORT NO ISBN-0-90169954-3
PUB DATE 80
NOTE 84p.; For related documents, see HE 016 159-167. Produced with the aid of a grant from the University Grants Committee.
PUB TYPE Guides - Classroom Use - Guides (For Teachers) (052)

EDRS PRICE MF01/PC04 Plus Postage.
DESCRIPTORS Academic Standards; *Course Objectives; Educational Objectives; *Evaluation Methods; Foreign Countries; *Grading; Higher Education; *Student Evaluation; *Testing
IDENTIFIERS Great Britain

ABSTRACT

Educational assessment in higher education as seen by those who have a key part in the process is considered, with reference to the British system. Ways in which the assessment of students can be categorized, and the need for every assessor to have a clear and explicit purpose are addressed, along with the principles of fairness, validity, and reliability. It is proposed that departments should declare their position in respect to the different orders of knowledge, attitudes, and skills they assess and should define the levels at which these will be assessed. In other words, knowledge, attitudes, and skills are not to be thought of purely in cognitive terms. Attention is directed to the following modes by which students may be assessed: formal and informal assessment, formative and summative assessment, continuous and terminal assessment, course work assessment, assessment by examination, process and product assessment, internal and external assessment, and convergent and divergent assessment. In addition, the specific techniques (e.g., types of examinations, dissertations) used in assessment are examined, and marking and grading is discussed in regard to disputes about relative weightings, values, and systems. It is noted that assessors have both standards and external examiners to satisfy. (SW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

TEACHING IN HIGHER EDUCATION SERIES : 7

ED230104

Suggestions for the consideration of lecturers and others concerned with teaching in higher education

INSIDE ASSESSMENT

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

A.M. Ross

Univ of Lancaster

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

This document has been reproduced as received from the person or organization originating it.

Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

Barbara Cockburn
Alec Ross

Teaching in Higher Education Series
ISSN 0309-3565

No.7: Inside Assessment
ISBN 0-90169954-3
© School of Education, University of Lancaster

First published 1977
Reprinted 1978
Reprinted 1980

Published by School of Education, University of Lancaster,
Bailrigg, Lancaster, LA1 4YL, England.

Printed in England by Titus Wilson, Kendal.

FOREWORD

This booklet is one of a series produced with the aid of a grant from the University Grants Committee as part of its programme for the development of knowledge about and skill in teaching in higher education. The series is intended to bring together traditional wisdom (not a source to be neglected) research findings when available and relevant, and summaries of what is regarded by the informed as good practice.

The sub-title "Suggestions for the consideration of those concerned ..." echoes that of a famous manual for the guidance of elementary school teachers and this series shares the same purpose. The handbooks may at times be mildly hortatory but they are not meant to be prescriptive. Teaching is something of an art and the aim is to encourage, within sensible limits, something of a personal style. The "rules" (if such they be) are constantly broken by successful teachers but few, if any, get away with breaking all the rules all the time.

Universities and other institutions of higher education have traditionally accepted research and teaching as their principal functions. The systems developed for training in research are well worked out. Till now little has been done for teaching. The initiative taken by the University Grants Committee was widely welcomed and has led to an increasing awareness in higher education generally of the importance of giving to the teaching function the attention it deserves. This series is intended to offer help to those who wish to make an effort to improve their teaching. It is hoped that those who read these pages will, as a result, be better prepared for this important aspect of their daily work.

The authors would be glad to receive reports and illustrative material which readers may wish to send with a view to improving the next edition of these booklets.

C.F. CARTER
Vice-Chancellor
University of Lancaster.

CONTENTS

INTRODUCTION	i
PART I: SOME PRINCIPLES OF ASSESSMENT	1
1. What is assessment?	1
2. Clear purposes	3
3. Fairness, validity and reliability	4
PART II: WHAT ARE WE ASSESSING?	7
1. "Knowledge, attitudes and skills" - declare a position	7
2. The longer answer	10
PART III: MODES OF ASSESSMENT	13
PART IV: TECHNIQUES USED IN ASSESSMENT	18
1. The principal techniques	18
2. Examinations	19
3. Course work	26
4. Schemes of assessment	32
5. Question control	37
PART V: MARKING AND GRADING	41
1. Case law	41
2. Symbols	43
3. Standards	44
4. Marking scales	46
5. Combining marks	50
6. Compensation	52
7. Marking course work	53
8. Marking open book & prior notice examinations	55
9. Declaring a class or a mark for a unit	57
10. Classification for final award	58
11. The external examiner	66
12. Bias	67
13. Profiles	68
PART VI: CONCLUSION	71
FURTHER READING	72
REFERENCES	73

INTRODUCTION

This booklet is concerned with assessment as seen from the inside by those who have a key part to play in this significant part of higher education. There is a public aspect of assessment which is concerned with the affirmation of academic standards, with the declaration that certain individuals may fittingly undertake particular professional tasks and with the justification of the large sums now spent on higher education. These public aspects of higher education cannot be ignored by those who teach but this booklet is more concerned with assessment seen from the inside - with the practical problems which university teachers have to solve.

"Assessment" is the word we must use because the task is seen as something wider than marking, setting examination questions or grading. (Though these traditional practices do play a large part in assessment the new lecturer should, from the start, look for wider connotations.) The beginner's task is made more difficult because though much is written about assessment, practice in higher education seems to rely heavily upon lore. His first attempt at drafting an examination question may well be criticised as "not an honours question" or condemned as

being "unanswerable in the time". Experience within a particular department is the best guide to local practice but a discussion of some of the problems which arise can help the beginner to learn more quickly the subtleties of what has been called "the examination game".(1)

Let there be no doubt about the importance of the task; those who are responsible for assessment control the life-chances of those they assess and are responsible for the maintenance of academic standards. Assessment also has an influence upon student motivation and provides a means of two-way feedback between teacher and taught.

Teachers have to learn to adjust to the role of being those who assess. This change of role is, perhaps, seen most clearly when prior-notice examinations are used; once the questions are declared the teacher should no longer tutor for now he is the examiner. The duties associated with the one task conflict with those of the other. The helping attitudes appropriate to the one do not go well with the judicial tasks appropriate to the other role.

In British practice the external examiner is always there to help the internal examiner to maintain standards; obviously he or she can be particularly helpful in settling marginal cases but the system relies on the internal examiners themselves being conscious of the need to maintain appropriate standards. There are too many examples of devalued degrees to allow any reputable teacher to be unconcerned about the need to seek to be true to the best traditions of scholarship, of teaching

and of integrity.

In the pages which follow the topic is dealt with in the following sections. In Part I: "Some Principles of Assessment" we discuss the ways in which the assessment of students can be categorised, the need for every assessor to have clear and explicit purposes, and the principles of fairness, validity and reliability.

Part II: "What are we Assessing?" presents the case that departments should declare their position in respect of the different orders of knowledge, attitudes and skills they assess and should define the levels at which these will be assessed. Knowledge, attitudes and skills are not to be thought of purely in cognitive terms.

In Part III: "Modes of Assessment", in describing the variety of modes by which students may be assessed, we have tried to provide a framework within which the main methods, or techniques, of assessment are discussed in Part IV: "Techniques Used in Assessment".

Part V: "Marking and Grading" deals with what is often the most difficult part of the assessor's task; this is where disputes about relative weightings, values, systems and so forth come to the surface. The methods used in arriving at final degree awards may seem pragmatic but we are always guided by principles and precedents. Assessors have both standards and external examiners to satisfy.

In Part VI: "Conclusion" we sum up.

PART I

SOME PRINCIPLES OF ASSESSMENT

1. WHAT IS ASSESSMENT?

Let us first be clear about what we are dealing with and what we are excluding. There is, as might be expected, a not inconsiderable literature on assessment; key phrases are student assessment, course assessment, student assessment of teaching, objective testing, evaluation, examinations, continuous assessment, course work assessment, terminal assessment, etc. etc.

N.D.C. Harris[2] explains some of these terms and makes a clear distinction between (a) assessment of the student and (b) assessment of the course. Our concern is principally with the former though it is obvious that the results of the examinations at the end of a course will tell us something about the course the students have been offered as well as a good deal about each of the students tested in those examinations. Harris distinguishes four ways of assessing the student: comparative assessment enables us to discriminate between the performance of different students. We do this when we grade students into classes of degree and, since comparative grading is involved, statistical manipulations can be of importance. So also can the "rules" (often conventions) determining when and how the calculations are

made. Do you round? Do you round the separate scores which go to make a final score?

Diagnostic assessment is used as a check on progress so far, with a view to making decisions about the next stage of the student's programme of study. Remedial work, repeating the unit, or diversion to an alternative pathway may well take place as a result of such an assessment. In some universities, first year examinations are used in this way, the student being permitted to proceed but only in the subjects indicated by the pattern of performance produced.

Absolute assessment appears sometimes in professional degree courses. A physiotherapist must be able to give the treatments successfully, a teacher must be judged capable of communicating to children, a pharmacist must be ascertained as being able to get the mixtures right.

Predictive assessment is rarely used in higher education apart from the belief that, in general, high performance in academic study is a sound predictor of competence in a wide range of demanding posts. Predictive tests have been used with great success in selecting men for training as pilots and some use is also made of such tests in selecting students for advanced work in physical education.

The reader will at once realise that there are elements of all these types of assessment within the conventional systems of student assessment used in higher education. The fact that a I or a IIi degree gives access to opportunities for research is an indication of the

predictive value we place on degrees and is arrived at after an exercise in comparison. We use the results to predict capability for research though we may not set the examination with that in mind.

It is unlikely that the assessment of students will become a series of precise measures of particular qualities producing a "profile" of the many dimensions of the student's abilities. The assessments we use in higher education will remain multi-purpose, somewhat imprecise and infinitely adjustable to the great range of circumstances to be found in higher education.

2. CLEAR PURPOSES

That assessment in higher education does remain flexible and multi-purpose is reason enough for teachers to recognise how important it is carefully to consider just what is implied in their own assessment of students. They should be clear as to the particular purpose or purposes of the assessments they are making and should carry out what checks can be made to discover how effectively those purposes are achieved.

Thus in one faculty of engineering high performance on a preliminary course in mathematics did not, in the event, augur well for high performance in a more advanced mathematics course in the same faculty.[3] If good prediction was required how could that preliminary test be improved? What other purposes was that same test successfully achieving? These are the kind of questions which the informed teacher in higher education will not only ask but seek to answer.

One of the first principles of assessment, then, is that every assessor should define, as clearly as he can, the purposes of the assessment. Those who are using more sophisticated ways of developing their courses will relate the assessment to the objectives set out for the course as a whole. If the aim is to provide opportunities for "critical thinking", or to apply known techniques to unusual situations, then the assessment scheme should include opportunities to demonstrate ability in such activities. The beginner has been appointed to his post principally because of his knowledge and skill in his subject; the content of his assessment schemes is likely to be thought through in terms of covering the subject. The subject matter is important but the assessor should keep in mind the need to judge how well the student has understood the concepts and mastered the skills and techniques which the course was designed to communicate. It is the student who is being assessed and the scheme of assessment should, therefore, be student rather than subject-centred.

3. FAIRNESS, VALIDITY AND RELIABILITY

Every student has the right to assume that his assessment is "fair". In this context fair can mean many different things quite apart from the obvious ones about security, cheating and equivalence, between students, of conditions, which need not be laboured; we can also leave aside such less obvious points as the importance in some forms of assessment of the ability to write quickly. Students have been classified by Miller and Parlett[1] as "cue-seekers" or as "cue-deaf"; certainly those who teach must be careful that they do not inadvertently provide

messages which only those with highly-tuned receivers can hear.

Standard texts such as Thyne[4] define as two of the desiderata for a fair examination, its validity and its reliability. Validity has been implied already since it is a measure of how well an examination does what it purports to do. Reliability is a measure of the stability of any test; it indicates how consistently a test produces the same result from one occasion to another.

These two principles derive from the statistical study of tests which are logically constructed and whose results can therefore be analysed and manipulated with mathematical rigour. Obviously standardised objective tests can be developed to the point at which they are, for example, highly reliable (see No. 9 in this series: "The Use of Objective Tests"); it is more difficult to find ways of achieving such a desirable end in the case of traditional university assessment schemes.[5] This does not mean we can ignore these principles; the onus remains on every examiner to ensure that there is as high as possible a measure of consistency in his assessments.

Traditional three-hour papers contributing to the class of a degree should always, in the case of subjective answers, be marked "blind". (This means double-marking with the second marker unaware of the first marker's judgment.) Significant differences can then be discussed before an agreed mark is given to the external examiner. Since marking is often done under pressure of time this check is doubly necessary.

Additions should also be checked; administrators often comment that academics argue about three marks having failed to notice addition errors of five. Inconsistencies between markers will be discussed further below. Inconsistencies in the way students respond can, to some extent, be reduced by care in the framing of questions and assignments. The formulation of any question should always reflect the purpose behind it and every question should be broadly representative of the field studied. When a large class has been split into several tutorial groups taught by different tutors then it is obvious that an examination paper for that course should reflect the work done in all the groups.

A scheme of assessment should therefore have a clear purpose, do what it sets out to do and do it efficiently, fairly and expeditiously. This is easily said and written but, as the beginner's first experience of acting as an assessor will surely demonstrate, the achievement of these simple objectives is not accomplished without considerable effort, ingenuity and hard work. We turn now to some of the finer detail in assessment.

PART II

WHAT ARE WE ASSESSING?

1. "KNOWLEDGE, ATTITUDES AND SKILLS" - DECLARE A POSITION

It has already been suggested that to improve our work as assessors we should begin by clearing our minds about what the courses we teach are intended to achieve. It is obvious that this should be defined not in terms of what the teacher teaches but in terms of what the students are enabled to learn.

Many commentators have suggested that the task of specifying our purposes - setting out our broad general aims as more precise objectives - can be assisted if we consider the work of Bloom, his associates and followers, especially the schedules given in the Taxonomy of Educational Objectives.(6) Bloom's list acts as a check-list which enables the teacher to consider the levels of intellectual demand being made upon the student. Bloom's cognitive domain includes, for example, the following categories:

1) Knowledge. This refers to specific facts, terminology, definitions, but also to ways of handling knowledge such as identifying trends, clusters and sequences, classifying, establishing categories, making satisfactory generalisations, evolving theories.

2) Comprehension. This refers not only to understanding material but translating that knowledge into other kinds of statement, interpreting it and extrapolating from it.
3) Application. Can the student apply what he has learned to practical issues and problems? Can he apply it to unusual situations?
4) Analysis. Can the student analyse what has been presented, identifying the elements within it, the relationships between these elements and the pattern of their organisation?
5) Synthesis. This deals with bringing points together to form a coherent whole. Can the student organise his knowledge to produce a workable plan for an investigation or a convincing essay on a problem?
6) Evaluation. Can the student make sound judgments and can he do this using external criteria as well as those which are internal to the problem presented?

It is clear that Bloom's checklist is arranged in increasing order of difficulty. Look at the questions set in examinations as the items in an objectives list: at what level in the taxonomy does one have to operate to answer these questions satisfactorily? And within each category there is also gradation of difficulty. Billing,[7] in a valuable application of the Bloom taxonomy to science education, instances the shift from knowledge of specific facts (What is the most important characteristic of a base?) to knowledge of theories and structures (What evidence best supports the theory of crystal-field splitting?). Clearly an honours degree assessment which remains at the lower levels of the learning-difficulty hierarchy cannot be satisfactory;

equally clearly a paper in which students could avoid all questions of a more demanding nature would be badly designed and could lead to injustice.

Bloom's cognitive domain is one of three which he hypothesised; the others are the affective[8] and the psychomotor and these too can be broken down into a hierarchy. The affective domain can, for example, be seen as moving from mere awareness of stimuli, aesthetic form and pattern through responding, valuing, conceptualisation to the development of internally consistent value systems. In the same way the psycho-motor domain permits identifications of simple perceptual skills leading on through physical skills to the smooth performance of highly co-ordinated complex skills.

Many of those who concern themselves with the quality of their work as teachers and assessors find the checklists of Bloom and his followers helpful whilst recognising that education - especially higher education - cannot be split into discrete domains and categories within those domains. The final year student violinist at the Royal Northern College of Music playing the Kreutzer Sonata is being tested at different levels of all domains simultaneously and much university work is of that kind. There is aesthetic appreciation in mathematics, and a cognitive element in high level gymnastics, literature requires awarenesses in the affective order as well as analysis of the kind found in the cognitive domain, and he is a poor scientist who has not developed an understanding of value systems.

"Knowledge", "attitudes" and "skills" as they are some-

times called are components of the learning the students acquire. Differences of view about the merit of a particular answer may well be explained by the differing weights given to these components by the examiners. Students following multi-subject degree courses are quick to "suss out" (as they put it) the assumptions built into each department's expectations. Answers incorporating an emphasis upon affective order outcomes may be more highly valued in one department than in another department where the cognitive domain is the only one which matters. There is much to be said for such diversity of practice; there is little to be said for not declaring the departmental position; there is nothing to be said for not being aware of the existence of what could be described as bias.

2. THE LONGER ANSWER

The short answer to the question "What are we assessing?" is "We are assessing the student's acquisition of the knowledge, attitudes and skills specified (or implied) in the objectives of the course followed." The longer answer would have to specify the levels to be tested and this would, of course, be related to the standard of the course.

The relationship to the objectives of the course is important; some authorities, anxious to keep the emphasis upon student-learning (rather than teacher-teaching), ask for these objectives to be stated in terms of the behaviours which the student is intended to display. What will the student be able to do? The statement of

objectives then includes verbs such as

to calculate
to explain
to illustrate
to predict
to judge.[9]

This point is picked up in the setting of traditional examination questions; it is obvious that a question beginning "Describe ..." is less demanding than one beginning "Distinguish ...".

Much used initial words such as "Discuss" (with "critically" added to indicate what kind of discussion is looked for), "Evaluate", "Assess", "Compare", "Contrast", "Explain" etc. are used to indicate the level of discourse which is sought and this, in turn, reflects what was intended to be learned. There is a growing literature on objectives in higher education; the following list of objectives for a mechanical engineering course is taken from the extremely useful book by Beard, Healey and Holloway.[10]

"a) Students will be able to recall from memory:
   1 the circumstances and qualitative features of the main tbl phenomena of thermal power and process engineering (flows in pipes and annuli; flow around aerofoils; film-cooling flows; free turbulent mixing layers, jets and wakes).
   2 the distinction between a tbl and a turbulent separated flow, and the circumstances giving rise to each.

b) Students will be able to derive:
   1 velocity and temperature profiles in Couette flows from prescribed transport-property assumptions, when the density is uniform.

c) Students will be able to determine:

1 the basic assumptions and the computational procedure of methods of tbl prediction which they encounter in the literature."

When course objectives have been expressed as clearly as this the task of assessment is made easier; the link between objectives and assessment is evident. If the objectives have not been fully articulated, the discussion arising from the assessment process may well lead to a clarification of the course objectives.

PART III

MODES OF ASSESSMENT

In theory the assessor should match the mode of assessment to the objectives or - as the jargon has it - to the "intended learning outcomes". In practice there is a framework for the institution within which methods must be found and in any case the beginner has to accept what is established departmental practice. Nevertheless there is much more variety in the system than universities, in particular, are given credit for; a conference on "The Assessment of Undergraduate Performance" held in 1969[11] heard papers from Professor Brockbank of York on innovations in examining arts students, Professor Page of Bradford on experiments in assessing applied science students, and Professor Elton of Surrey on new techniques in the field of pure science. These papers, together with an overview of research by Professor Beard (now of Bradford) summarise the developments of the 1960s. The 1970s have been no less fertile in the production of alternatives to the still dominant traditional three hour unseen paper. There is no lack of experiment and the following pages are intended to provide a structure within which the main methods may be discussed.

Rowntree[12] distinguishes the following modes of assessment:-

Formal and Informal Assessment. Set pieces which "count" are to be distinguished from the appraisal of things the student may be doing anyway. Rowntree illustrates the latter beautifully with a remark taken via Bourdieu and Passeron[13] from Stendhal to the effect that a seminarian's progress in piety may be judged from the way he eats a boiled egg. Students may justly claim that they should know what does and what does not "count". In professional courses (medicine, teaching) such informal assessments may not contribute marks but can influence markedly what is written in confidential references.

Formative and Summative Assessment. Formative assessments are given during a course, being designed to provide feedback to help in shaping a student's future learning; summative assessments aim at an overall statement, usually at the end of a course, of where the student has got to. The categories are not, however, mutually exclusive; the assessment of a mid-course essay has a pedagogic function but may also contribute a small amount to the final summative assessment.

Continuous and Terminal Assessment. These modes of assessment are now very familiar. Some universities and colleges dislike the connotations of continuous assessment and use the term "course work assessment" instead. Students in the 1960's seemed to prefer continuous assessment,[14] especially when combined with terminal assessment, but some object to the constant pressure which the system implies and the present state of student preference is unknown. Though terminal assessment often means traditional examinations it need not do so; furthermore continuous assessment can include formal examinations.

Course Work Assessment and Assessment by Examination. The examination mode is familiar; course work is less clear and when it is used as part of the assessment scheme the limits should be defined. Both examinations and course work of different kinds are fully described in Part IV and the problems associated with marking course work and with marrying course work and examinations within one assessment scheme are given special consideration in Part V.

There are powerful arguments against examinations; there are also arguments to be advanced against course work assessment. It is rumoured that there has been at least one student protest march behind a banner demanding "Give us back the freedom of our examinations".

Process and Product Assessment. We have tried in these booklets to avoid using the technical terms (jargon?) of the professional educationists but technical terms are often useful when one has to categorise. We spare the reader the contrast between idiographic and nomothetic assessment though we wish to mention in passing that Rowntree (ibid p.158 ff.) deals with this particularly well. "Process and Product" is, however, a pairing well worth scrutiny.

Traditional assessment examines where the student has got to, that is the product of learning. There in one place is the examination script, the dissertation, the project or the exhibition. The success of the teaching is judged by the results of the learning rather than by the evidence of the process of learning. Yet assessment of the process (how he learns, how he gets to his

solutions) may well tell us more about the student than the product does. This approach may appear in some intermediate examinations of the formal type (for example, Part I examinations) and can lead to an assessment saying "This candidate has written a poor script but the quality displayed in his first question is such that I am convinced that he is beginning to grasp the concepts needed to develop an understanding of this subject." Such an adjustment in medias res would not, of course, be possible at the final assessment.

Internal and External Assessment. This again is familiar; the teachers are wholly responsible for formative assessment but when it comes to summative assessment and above all to classification an outside assessor is brought in. The final assessment is technically that of the internal assessors moderated by the external examiner. Moderation consists of making the marks consistent within and between papers for work of a similar standard but also of bringing the classes represented by those marks into line with standards elsewhere. The external examiner's judgment is usually, by regulation, final but even if it is not he can always use the ultimate deterrent. "You can give him a first if you like but he would never get a first in my university."

Convergent and Divergent Assessment. The familiar distinction is between styles of thinking and of expression, that is between the convergers, those who excel when the problem has one clear-cut answer, and the divergers, who prefer open-ended problems which allow them to use their own free-ranging ideas, their creativity and their imagination to the full. The objective test (see No. 9 in this

series, "The Use of Objective Tests") is clearly a useful way of assessing what might be called "convergent material" but might, at the same time, prevent the divergent student from displaying his particular qualities. Similarly an examination question of the "If Cleopatra's nose had been half an inch longer ..." type could be thought to discriminate against the converger.

These then are some of the modes we may use in assessing students in higher education; few departments or institutions will rely on any one type. Indeed the picture we see when we look around us is a confusing mixture of all the elements we have listed, together with several more which we can only describe as "local peculiars" defended, of course, by the hallowed non-argument "But we have always done so."

The assessor should be aware of the relationship between the learning intended and the mode of assessment chosen. He should try to determine what the strengths and weaknesses of each mode are when set against the context of his students, his subject, his department, his college or university and then, within the limits set by the institutional framework, he can find a pattern of assessment which best suits his purposes.

PART IV

TECHNIQUES USED IN ASSESSMENT

1. THE PRINCIPAL TECHNIQUES

In this part of the booklet we shall take for further scrutiny the principal techniques used in undergraduate assessment and look at them in two categories: traditional examinations and course work assessment. Two important techniques of examining which may be used in either examinations or in course work assessment - objective testing and essays - are dealt with separately and fully in two companion booklets in this series: "The Use of Objective Tests" and "Essays".

Examinations include:

* traditional three hour unseen essay examinations;
* modified essay examinations called short answer examinations, where students write shorter answers to a larger number of more precise questions;
* open book essay examinations during which students have access to specified texts;
* prior-notice essay examinations where questions are given out for students to prepare in their own time for a stated period before the examination;
* objective test examinations;
* oral and practical examinations.

Course work includes:

* essays;
* long essays;
* dissertations and theses;
* reports on laboratory or field work;
* practical work such as laboratory exercises, translations, musical composition, clinical or seminar performance;
* projects;
* short tests.

2. EXAMINATIONS

We begin with the traditional examination and it is salutary to quote Rowntree's list of features.[12]

"1) All students are given the same task to perform.*
2) Although students are expected to have 'prepared' for the task, its precise nature is not made clear until the moment they are asked to embark upon it.
3) The task is performed in the presence of a non-collaborating person with the policing role of 'invigilator'.
4) All students are given the same limited period of time to work on the task even where the task is one which they might be expected to perform better if given no time limit - as they might well be with similar tasks outside the examination situation.

---

* In fact students do not all tackle the same task; there is a choice of questions and if these vary in difficulty or in scope this poses problems for the assessor.

6) Students (or candidates as they are often called in this situation though they would not be when performing similar tasks as part of course work) are not allowed to consult with one another.
7) Students are expected to experience at least some sense of stress and urgency while performing the task - again regardless of whether such would necessarily accompany their performance of similar tasks elsewhere."

The strength of the formal examination lies in its success in controlling many important factors which could unfairly influence results. Nevertheless it could be said to favour those who work well under such constraints and to put at a disadvantage those who are disorientated by stress.

Lovell[15] provides a useful summary of the criticisms of examinations and mentions

* their distorting effect on the curriculum,
* the values they reinforce (e.g. competition) may not be desirable,
* the probability that some candidates do not do themselves justice,
* the cramming they engender,
* their relatively poor predictive value which may in some cases be due to inadequate sampling of the candidate's knowledge (compounded, no doubt, by successful question spotting),
* inconsistencies between markers.

Nevertheless much can be done to improve examinations; apart from the point about clarifying purposes, already made, there can be a full discussion of the questions

and of what they are designed to produce, there can be an agreed marking scale and a determination to ensure that choice of questions, other things being equal, should not determine placement in the rank order.

Short answer essay examinations

This variant of the traditional examination is designed to correct two drawbacks: inconsistent marking and inadequate coverage of course content. The candidates are required to answer a greater number of more precise questions, probably in the usual three hours. The questions must be considerably increased in number beyond the usual three or four, however, if sampling is to be significantly improved. Raising the number to six will make little difference.[16] Perhaps 12-20 questions would be a reasonable number. This would be equivalent to 3 or 4 "gobbet" questions in a traditional examination, that is questions requiring students to comment on a small number of short passages, quotations or statements.

Short answer essays have a special part to play where a wide field of knowledge must be examined but where it is desirable, in the eyes of the examiners, for candidates to be asked to organise and to present the knowledge they are asked to recall in a purposeful and lucid way. The candidates will not, of course, have the opportunity to develop answers in the way that a traditional examination question allows them to do.

A special case of the short answer essay question is a Modified Essay Question developed in medical schools where a series of such questions forms a coherent step by step examination of the several aspects of a complex problem. In medicine this is a means of assessing candidates' ability

to solve, for example, patient management problems. "The technique is in effect a series of open-ended questions about a developing situation in which the responses at each stage are relatively free."[19]

When an essay is broken down into several short answers marking consistency noticeably improves.[20] The short answer examination allows the examiner to ask more precise questions and thus to construct a more detailed marking scheme.

Open book essay examinations

This variant reduces the burden of memorisation; students are given access to specified texts during an examination which is usually traditional in form but may be a short answer essay or objective test examination. (The Modified Essay Question discussed above might be used in conjunction with reference material as a source of background information.)

The "open book" could be a set of duplicated notes or papers, specified chapters from a textbook; a poet's works, even a dictionary, fall into this category. More wide-ranging standard works of reference may also be permitted. "Candidates may find the following texts of use; they may be brought into the examination, but only in the editions specified." Often the texts are supplied or deposited the day before to be checked to see that they are "clean". The use of open books has not apparently led to any improvement in overall academic performance.[21] There may be an interactive effect with the results of particular papers.[22] The questions can be more difficult to set than traditional essay questions on the same topic since they have to seek out aspects of the problem not dealt with directly in the

text given to the examinees.[21]

Experience shows that students need to be trained to use open book examinations to their best advantage.[23, 24] Some, knowing that texts will be available in the examination room, arrive there ill-prepared and waste time during the examination. On the whole students react favourably to open book examinations[25] though a student already under stress may well find such an examination more hindrance than help.[23]

Prior notice essay examinations

In this case students are given notice some 10 or 14 days before the examination, or the date for submission, of what is to be examined. Variations include

* giving the exact words of the question,
* indicating the general areas,
* asking the student to turn up on the appointed day and to write his answer under examination conditions,
* permitting him to write his answer in his own time and to present the finished answer by the appointed time having had full access to reference materials as he writes.

When the student has to write answers, under examination conditions, to known questions, the examination may become "a dull test of memory or a writing exercise"[23] as the students try to get down their prepared answers. Anxiety does not decrease[26] and they tend to over-prepare.[24] The answer is to give prior notice only of the general area to be examined and to give out the precise wording of the questions only when students are in

the examination room. (They may be allowed to bring in notes and texts as required.)

There are practical problems to consider. For example, during the period of preparation key texts must be kept available to all students, including others who may be being assessed conventionally. Tutoring obviously stops but other sources of help need to be controlled.

Practical examinations

The same arguments can be applied to practical examinations. Practical skills are an important part of many courses in higher education and in some cases they are underweighted in assessments. This may be due to the fact that they are often simple performance skills which every student must have but which do not discriminate. It is perhaps unfortunate that there is an increasing tendency to assume that everything which is presented for assessment must necessarily "count" towards the final assessment even when - as with some of these skills - they could be formally assessed at different points in a course to ensure that a satisfactory level of mastery has been achieved long before final assessment takes place. This skills assessment could then be of no further relevance and would have no place in final assessment.

Other more complex and more important professional practical skills (in medicine, engineering or agricultural sciences, for example) may be formally assessed as part of final assessment. Bligh[27] reviews the literature on practical examinations of this kind; Fleming et al.[28] discuss their particular use in medicine.

Objective test examinations

Objective tests are the antithesis of essays. The neatness of the antithesis has led to many published expositions of the opposed characteristics of the two methods.

Objective tests require candidates to answer a large number of compulsory questions to which they must make fixed responses such as choosing one possible answer from five. Objective tests are uniform for all candidates; essay examinations allow them a choice - though all may sit the same paper the examiner does not mark the same papers. Objective tests are marked quickly, accurately and consistently; essay marking is notoriously unreliable. Essay questions are frequently wide enough for students to take them in several ways and their responses are free. However, fewer essays can be written so while objective tests allow the examiner to sample performance over a wide field of knowledge essay examinations do not. Clearly the two methods are complementary.

Objective tests can be used for examining higher order abilities but the questions are always difficult to set and higher order tests are more complex, more open to error and more dependent on the expertise of the test-setter. They are probably most effectively used at more elementary levels and for more factual subjects.(29) "To spend too much time devising ingenious Multiple Choice Questions to test understanding is as wasteful as using vivas only to test factual knowledge".(30)

Objective tests are traditionally given less weight in overall assessments than essay type examinations, but they are used extensively to measure mastery of basic

knowledge, or of the core of a curriculum. The virtues of the objective test as a precise instrument of measurement should be recognised more widely. Fixed response questions are not to be thought of as trivial. Their positive advantages are that each test has a single clear purpose, each question tests a discrete item of knowledge or a discrete skill yet the questions relate logically to each other.(31)

3. COURSE WORK

Course work may occasionally be the only form in which students' work is assessed; more commonly it complements examinations. Course work assessment may be capable of measuring what formal examinations claim to measure but at different times (providing a better sample of students' performances) or in different ways (providing a better general picture of overall attainment). It may also be capable of measuring what formal examinations cannot accomplish and, since less weight traditionally attaches to course work (in relation to the amount of work submitted), the examiner has room for experimenting with unusual modes of assessment.

The inclusion of an element of course work assessment in an assessment scheme does take some of the pressure off the bad examinee and gives him the opportunity to build up the evidence needed for him to satisfy the examiners of his abilities. Of course, spreading the load of assessment may spread the anxiety too. This is a factor advocates of course work assessment should not ignore.

It is not possible to generalise about students' attitudes towards course work assessment with confidence. The evidence we have is mainly from broad surveys of student opinion;[14] we need to know more about how they respond to the volume, nature and use of course work assessment in different kinds of courses. In the 1960's (as has been mentioned) there seemed to be a general preference for course work assessment as opposed to traditional examinations.[14] This may not be the case today. The most enthusiastic calls for more course work assessment may also have come from those students who had limited experience of this mode of assessment.[28]

Possibly, then, as course work assessment does increase in volume, so students begin to voice criticisms. The burden of essays, the subjectivity of marking, and the relation of course work marks to examination marks are all questioned. So, too, is the way in which a student who concentrates his studying on work for essays increasingly narrows his field, for he may be able to draw on essay material in examinations, especially when there is a wide choice of questions. The student who has been reading widely may feel at a disadvantage.[32]

Course work assessment suffers all the disadvantages of the less conventional examinations: plagiarism; unequal availability of resources; too much formal work for assessment leaving students less time for independent study; difficulty in enforcing or even establishing uniform procedures (students handing work in late may gain unfair advantages); emphasis on consistency of effort and performance rather than on final achievement (which may penalise the student who spurts at the end

of a course); and inconsistent marking.

The loose framework within which course work assessment is conducted, the paucity of research into the effectiveness of different methods and the infrequency with which either course work assignments or course work marking is scrutinised make it difficult to use systematically.

Essays

A great deal of course work is in essay form. Assignments vary considerably in their length, less so in their form. They vary from a normal essay of about 2,000 words, to long essays of 2,500 - 4,500 words, to dissertations of 5,000 - 15,000 words.(33) The full range is seldom used simultaneously; indeed there may be tension between various types. Where dissertations are used they are given greater weight than ordinary essays; where there is systematic use of essays dissertations are rare.(33) Dissertations may be used to replace examinations; essays are normally not so used. A dissertation is in every respect a more substantial piece of work than an essay.

Formal assessment of essays always includes writing skills but style, elegance and economy of expression may sometimes overshadow other important qualities; the subjectivity and inconsistency of marking is their main drawback.

Essays are worked into a formal assessment scheme in various ways. Students may be required to write a number of essays all of which are formally assessed; they may be permitted formally to submit, say, only the best 6 out of 8 required essays which may be selected by staff

or students;[34] sometimes they may rework essays originally presented and discussed as seminar papers, re-writing them with tutorial guidance before submitting them for formal assessment.[35] On the other hand essays may be moderated by seminar performance.[36] Essays may be used in conjunction with final examinations as a 14-day test.

A 14-day test (involving the writing of one or more essays) is closely akin to prior notice examinations. Beard[24] feels that the time constraint brings this kind of examination closer to real life situations. Given more time than he has in an examination the student is tested on his ability to establish priorities and to structure his material, given less space (such essays may be required to be of no more than 1,200 to 1,500 words) he has to organise his material more effectively and to concentrate both his thoughts and his expression. He may have a better opportunity to develop his own ideas but higher standards are usually expected than is the case for examination answers.

Extended essays tend to be used for the writing up of small scale projects. They fall between standard essays and dissertations, requiring more work than the one but falling substantially short of the kind of investigation a dissertation requires.

Dissertations

What is there that is distinctive about a dissertation? Is it different in kind from a very long essay? Students who do well on course work dissertations may not be those who do well in traditional examinations;[37] perhaps a

good dissertation requires the use of different kinds of ability or it may be that some students give of their best when they have freedom of time and space in which to work. Students who opt for dissertations are often those who are involved with and interested in their work; working on a dissertation may also kindle greater interest. "Great originality is neither required nor expected; the essential task is to unfold a coherent and progressive argument leading to a solution of the problem set or to significant conclusions to an investigation."(38) But the opportunity for creative work is there; a dissertation gives students opportunities for creative work in almost any subject. A dissertation allows them to set out what they have learned in a more complex, thorough, elaborate or subtle way than they have room for in an ordinary essay; it is a study in depth.

Dissertations take up a great deal of time (they are often vacation work); they may lead to students over-specialising in too limited a field; they are difficult to integrate into a course.(23) Some institutions limit the number of dissertations which a student can offer as part of his terminal assessment.

Projects and Practicals

As every quizzical observer of practical course work sooner or later notes, what is actually examined is often a report and not the practical work itself. To students results are what counts, together with their interpretation, theoretical importance and presentation; they rapidly familiarise themselves with all the sources of 'good' results. Perhaps it is not surprising then

that, if a recent report is anything to go by, many practising scientists admit to 'fudging' results at least once in their careers.[39]

The abilities required for good report-writing are not, of course, in themselves of negligible importance but perhaps some effort should more often be made to distinguish between and separately to assess report-writing abilities and practical abilities. Short practical tests or multiple choice tests might sometimes be used for assessing both knowledge of techniques or procedures and the logical choice of steps to take in an experiment or exercise, and also the ability to carry them out in practice. Skills which later may be the essential components of higher order practical abilities can be discretely examined in this way - perhaps on a pass/fail basis.

Mastery tests for practical skills may be one area where a pass/fail assessment is more appropriate than classifying marks; it is often suggested that it may also be so for projects. Projects, like dissertations, present students with problems or investigations whose solutions or results cannot be found ready-packaged in the library, and sometimes too the opportunity to do original work in a narrow field. They are an extended piece of work, often carried out by small groups, and hopefully do not duplicate work performed by others; practical work may actually be the best source of results.

An important element in project assessment may be its contractual aspect;[27,40] students and examiner discuss what can and will be assessed before they begin.

As the work progresses they may then present draft reports for discussion and criticism, being free to incorporate in the next stage of the work any practical points that emerge. Pass/fail assessment might be used to allow examiners simply to judge whether every aspect of the project has been successfully handled - whatever its outcome. Assessment is necessarily subjective and more than one examiner should be involved.[29] Project work and project assessment have been described and discussed in recent publications.[40,41]

4. SCHEMES OF ASSESSMENT

There is, then, no lack of variety in the techniques available to the assessor and there is a good deal of experimentation in developing new ways of assessing. The new universities in Britain took a lead in introducing a course work element, usually as something put alongside the examination mark derived from the traditional three hour test. The newly appointed tutor will usually find that there is a norm in his or her department; this is the scheme whose complexities he must set himself to master before ever he begins to think about new methods. New methods and new approaches to assessment cannot, however, be introduced unless there is a reasoned case for doing so.

When there is such a case then the choice of modes and techniques of assessment will obviously be that which best fits the particular task of assessing what the course in question was intended to produce in the students. Specimen assessment schedules must then be prepared and discussed. Questions such as that of the

weighting of the different assessments included in the schedules must be resolved. Second year course work assessments may be given less weight, for example, than final year assessments; examinations could provide, say, 60% of the final assessment while course work made up the remaining 40%. Then there are questions such as whether the examinations will cover the whole area of the course or only those parts not covered by course work assignments. Should students who have done a dissertation in a particular area be prevented from answering examination questions on topics from this area? How can we know whether in a large course there are going to be instances of "undue overlap" - and what is to be done if examples of this do occur?

The drift of our argument is that a mixture of assessment techniques is likely to produce a better picture or profile of the student's abilities. A paper by Heywood[42] shows how in four types of assessment - objective tests, essay-type examinations, term-time essays and tutorial assessments - variable factors might possibly influence performance and do so differently according to the type of assessment used. The variables he mentions are

* personality factors such as motivation, study habits, extraversion and emotional stability,
* extraneous factors such as working speed, examination technique, ability to write intelligently and fluently, and physical well-being on the day of the examination,
* chance factors such as sampling errors on the examination paper, subjective bias in the assessment and errors in combining marks.

A recent study by Entwistle and Wilson[43] gives ample evidence of how all the above factors can influence the result of an assessment. In such a situation it is clearly wise to consider introducing a range of assessment procedures when this is consistent with the objectives of the course. This argument would be particularly strong if different techniques produced quite different results; thus in one study students who did well on projects were not those who did well in unseen examinations.[37] We hesitate to ask "Which was 'right'?"; the prudent course would be to keep both types of assessment in the scheme.

The traditional examination is not without its critics[44] but it dominates the field and is likely to continue to do so. Our task is then to improve it, to use it where it is most efficient and to use other methods elsewhere. Course work marks tend to be higher than examination marks; this does not mean that course work marks give the truer picture of the student's relative performance. Sometimes, as we have seen, course work marks are made up of the best pieces of course work, occasionally an essay is given therapeutic, that is generous, marks for sound pedagogical reasons; sometimes second year course work is marked less rigorously than third year or final examination work. Obviously good students will tend to do well in both course work and examinations; indeed outstanding students will shine whatever mode of assessment is adopted.

But we are, of course, concerned that our assessments shall give as true a picture of performance as is possible for all students, not only for the good students.

Brockbank is of the impression that only a small number of students gain noticeably different marks when assessed one way rather than another and while a mixture of techniques may occasionally (because of the nature of the mix) produce rough justice nevertheless this is probably the fairest approach to assessment.[34] In his department, where a fairly wide variety of methods of assessment was used, he thought about 65% of students obtained comparable grades over the whole range of methods, 15% did better using methods which allowed them to work at their own pace and 15% did better when they were examined under time constraints.

In constructing a mixed methods assessment scheme, then, balances must be struck. (Examples of such schemes can be found in the Nuffield Newsletters.[45]) Not only must the virtues of the modes and techniques of assessment available to us be matched against our purposes, and the obvious merits of different techniques weighed against each other, but their disadvantages must also be taken into account as we have just suggested.

Consider again the question of the validity of a technique of assessment; we always seek the highest possible validity for our assessments. Validity is always high when questions can be framed as precise "performance directives" which allow the assessor to predetermine marking criteria[4] but this is an aspect of validity which we ignore when we use essay-type assessments whose questions are designed to stimulate the production of students' own thinking on a problem and which give them freedom of response. How many questions should take this form? We should ask whether the essay-type

examination papers we set should be structured so that every student answers a specified number of questions of this type and specified numbers of other distinctively different types of question.

Validity is also high when the questions asked are a good representative sample of the questions which could be asked.[4] In an essay-type examination which does not allow wide and detailed sampling of syllabus content should every question be broad and comprehensive in its scope? Or should this type of examination be reserved for any kind of question which fully utilises the strengths of the essay as a technique while objective tests are then included in the scheme to compensate for the sampling deficiencies of the essay examinations?

Validity can also be described as curriculum validity (see "The Use of Objective Tests"); the methods used for assessment should reflect students' learning experiences during a course. Thus projects which were a first experience of working independently would not normally be formally assessed and students whose course has emphasised the discussion of issues would not normally be examined by a factual objective test.

Practical problems inherent in the use of particular techniques must always weigh heavily with assessors. Is a method easy to use, does it require special facilities, does it need rehearsal, will suitable questions or assignments be easy to find? Good new project assignments are often hard to come by when the method has been used in a department for a number of years, for example.

Different methods may also impose widely different workloads on the students. Thus the preparatory work a student can do for practical examinations may be limited by the availability of laboratory time, equipment and materials; the labour which can go into course work essays may be limited only by the hours a student is willing to work. Prior notice examination questions may make unfair demands on a student's time when they are set at the same time as unseen examinations. Course work assignments must be set with an eye on what students are being asked to do on their other courses.

Finally the students themselves have a point of view which should be recognised. Their prejudices against,[23] liking for[14] and expertise in particular methods[29] should be borne in mind so that the system of assessment is seen by all the partners as a reasonable solution to an extremely complex problem.

5. QUESTION CONTROL

It is sometimes suggested that closer control should be exercised over the actual questions and assignments used in formal assessments. It is not easy to find out how widely this is done nor to take up a position on whether it should be; we offer the following points for consideration.

Apart from the nature of its content and its intellectually pleasing aspects, the virtues of a "good" question are those which enhance its validity. They include:

* relevance to the purposes of an assessment;
* adequate content, giving the students scope

fully to demonstrate the knowledge, attitudes and skills being assessed;

* an appropriate level of difficulty, being within every student's reach, appropriate to the level at which they are being assessed, but producing a spread of scores;
* being suited to the time available; every student should be able properly to attempt each question - more time is need for use of knowledge than for recall, still more is needed if they are expected to think deeply about a problem;
* clear instructions.

The case for more systematic control of questions rests, for the most part, on analyses of traditional examinations, particularly in the sciences, which showed that a majority of questions required little more than knowledge and comprehension of a topic and made few of Bloom's higher order intellectual demands on the candidates.[46,47,48] The inference is that in the absence of any detailed specification of what is to be examined assessors with a feeling for exactness may tend to ask more and more factual questions. (But this is not at all typical of non-science examinations which sometimes make excessive demands on higher order abilities.) Black[47] recommended that departments should ask examiners to present an analysis of the abilities they claim to be testing in relation to their actual questions, together with model answers, where appropriate, and with detailed marking schemes.

Where there is a choice of questions, as has been noted, the equivalence of questions may be ignored. Departments may have a role to play here too, not only in

deciding how important the element of chance or skill in choosing questions should be, that is of insisting on equivalence if it is desired, but also in controlling the quality of individual questions. When questions are unequal the style of a question may be an important factor in a student's choice of questions to answer and thus of determining his success.

Departments often have conventions; should they formulate and publish rules about the style of questions?[47] Less able students, according to the "July Thoughts" of two mathematicians, may be led astray by the style of questions.[49] They may be drawn to questions which imply success, and a narrow search for answers, such as those triggered by "Find", "Define" and "Prove". It is the "We've done it before" approach; in fact they may have been deceived and have not "done it before".

A systematic review of all the questions and assignments he sets may be an essential exercise for every assessor; while this is often done with examination questions it is not so common for course work. Ideally a review would be a group exercise. The content, difficulty, meaning and suitability of each question or assignment could be checked in turn. The justification "It's all right; they'll know what I mean" should not be accepted. Precise questions must not be trivial; comprehensive questions which require only a mass of reportage should be discarded. The second part of the review is to scrutinise the form and phrasing. What will the students take the question to mean? The new assessor could ask himself which phrases indicate an "easy" question in his discipline.

Finally, where are "good" questions to be found? Look through past papers. Find out what other universities include in their papers. Such questions are not always the product of systematic analysis of what is to be examined, nor, on the other hand, of the solitary imagination. Ideas for good questions are most likely to come to mind while teaching, preparing to teach or in the course of reading. Note them down at once; there are not so many of them that you can afford to let them pass. Having found them, however, remember that they do have to bear a direct relationship to your course and to the objectives, explicit or implicit, of that course. The good question relates directly to the content of the course, challenges the student by approaching the topic from a slightly unexpected angle but does this without excluding the candidate who is expected to appear in the lower half of the rank order. We should give ourselves time to write such questions and we should be proud of ourselves when we have produced one capable of eliciting the first class answer without by-passing the comprehension of those whose sights are lower.

PART V

MARKING AND GRADING

1. CASE LAW

In the introduction to this booklet we said that a good deal of day-to-day practice in the field of assessment is based on "lore". Lore is never to be dismissed out of hand; many a teacher in higher education has gone through the cycle of first dismissing traditional practice as being ridiculous, then having doubts when the "simple", "straightforward", "logical", "fair", "sensible", systems of their own devising turn out, when applied to real cases, to be capable of producing patently unjust results. Nowhere is this more true than in marking and grading.

The despised "lore" enshrines lessons learned from the experience of processing a very large number of real cases; it is case law and is not lightly to be ignored. Truth is rarely simple and natural justice may demand that in dealing with particular cases we have the room for maneouvre which allows justice to prevail. Students reading this are assured that the flexibility which exists in most systems is used in their favour; the powers resting in Boards of Examiners may be great but those powers are used, almost always, in a benign way.

Part of the reason for the reliance we place on our "case law" lies in the fact that institutions of higher education provide so many and such different kinds of courses and, in consequence, such a variety of types of assessable work. Often this diversity occurs within a single degree scheme. Now some work can be measured quite objectively; the marker has merely to check whether the student can meet the criterion or not - 100% can be given to the one who does it perfectly and 50% to the one who completes only half of the assignment. However most marking in higher education is less straightforward than this.

When there is no "correct" answer the marker has to decide how the candidate has performed in relation firstly to criteria which are not capable of precise definition and secondly to other candidates. A teacher who sought to teach the calculus would be happy if 90% of his class scored 90% or more in an examination which was designed to test competence in the use of the calculus at the end of the course. A teacher of constitutional history would be criticised if 90% of his class scored 90% or more. This obvious and widely recognised distinction is too often ignored in discussions about marking and grading.

The purpose of this booklet is not to offer solutions to the many problems which arise in assessment. We recognise that newly-appointed lecturers are involved, from the start, in putting marks on assignments, papers and projects which contribute to the student's final award. They are, therefore, working within the established practice of their department. However there is a continuing debate about assessment in most institutions of higher education; our intention here is to introduce

the problems surrounding what at first sight seems to be the simple task of marking.

2. SYMBOLS

We mark by putting on a piece of work a symbol (perhaps a tick or a cross), a letter (perhaps from another alphabet), a combination of letters and symbols (including ! ? and *), a number (sometimes in Arabic, sometimes Roman), a word (occasionally with the suffix "-ish"), a phrase or an extended note.

With short objective items (as in an objective test) only two marks are possible, right or wrong, tick or cross. With material which calls for discrimination between types of response, a wider range of marks is needed - we use a scale of marks and the first thing the new member of a department should do is to learn the scale which is in use and to discover the significance of the refinements which, by custom, have been developed. Teachers in higher education are able, ingenious, subtle and determined people; there is no end to the complexities which they can bring into the most simple system. Perhaps newcomers to certain departments should be given a licence to mark only after achieving not less than β/α (or possibly α/β) on a test which asks them to describe in not more than 500 words the quality indicated by the mark of β^{+} as compared with two of the following:- β^{-} β^{+-} β^{++}

Those who thus embellish simple marking schemes are merely stretching the scale to produce a greater number of intermediate points. For such people the initial letter is still an important signal, equivalent in the percentage

scale to the marks 50, 60 etc. Whether it be a letter or a number the mark is, however, nothing more than a symbol seeking to communicate a judgment about the quality or standard of the work assessed.

3. STANDARDS

Symbols matter not at all; standards matter a lot. For an objective item the question of standards is easily answered. "Can he do it or can't he?" The problem really arose at an earlier stage when the decision was made that this item of knowledge or of skill was suitable for inclusion in this course as an assessable component; this problem does not arise at the marking stage. The only problem remaining at the marking stage (and this would in fact have been dealt with when the test was devised) is the weight to be given to the correct response; is this a 1% or a 5% item?

The standard of the question is important in objective tests but it does not confront the examiner in quite the same way as it does when he tackles the marking of an essay whose author is seeking (perhaps not too successfully) to differentiate between the political thought of Locke and that of Hobbes. It must be repeated that a symbol fits a judgment; we should never let the opposite occur. Some writers favour the analytical system of marking in which, by a pre-ordained plan, marks are given for developing particular points likely to occur in the answers. Obviously such systems allow for highly unusual,creative and unconventional answers but if the system produces a mark out of accord with the examiner's overall judgment of the piece, let the

mark be changed. An overall judgment is made against ideas about "first class marks", "work of a lower second class quality and low in that category" or, perhaps, "borderline between first and upper second".

When work completed during a course contributes significantly to the final assessment there is a particular point to be settled beforehand (by the department) about whether to mark at finals standard or at the standard appropriate to the stage reached in the course. A common way of dealing with this difficulty is to underweight the mid-course element in the final assessment.

Every judgment about standards should also take into account the conditions under which the work has been completed. This brings us back to considerations about course work as compared with examination work which we develop more fully below.

One point on every scale, whether it be a letter, a word or a number scale, is of particular importance: this is the point which indicates the pass/fail line. We have all of us smiled wryly at the critic who remarks that with a pass mark of 40% an engineer who gets more than half his work wrong still passes! A scale which is used to describe a range of performance from the excellent to the hopelessly inadequate has no need for a "pass" mark provided that everyone who should do so does understand what different points on the scale indicate. It is this thinking which has led bodies responsible for examining in schools to abandon a "pass" point and instead to describe different levels of performance.

The insertion of a "pass" point is, therefore, in a sense, an optional extra but for some time it is likely to stay with us in higher education. It does perform the useful function of indicating the point at which overall performance is so poor as to make it necessary to indicate to the outside world that the student in question cannot be regarded as competent in the field examined. Note that a further judgment creeps in here. "Is he fit to let loose on the world?". A degree is also to be thought of as a credential which the examiners are offering to the public. The point at which the student can be regarded as "passing" can, therefore, be shifted. Some assessors have even been accused of shifting their pass standards to match changes in employment opportunities.

The important question to settle is, therefore, that of standards. Undoubtedly there will be cases which are difficult to decide and it is these which should be examined carefully by the external examiner who brings to the Board of Examiners experience of standards elsewhere. The marker who is genuinely puzzled as to what kind of mark to give owes it to the student concerned to seek other opinions, those of colleagues and of the external examiner. We expand on the role of the external examiner later.

4. MARKING SCALES

We have already referred to the different kinds of verbal, literal, numerical and symbolic scales which markers have devised. Many systems use a combination of these, switching from one to another at different stages in the process, usually going into the numerical scales at the

point where separate marks have to be conflated, though there are also ingenious matrices which are designed to combine literal marks.

Whatever scale is used we should beware of the overall effect which its usage has. In an eleven point numerical scale going from 0 to 10 a marker might readily give 7 out of 10 when he might hesitate to give 70% on a percentage scale, especially if he knows that 70% means, by custom, a first class mark. Clearly it is necessary firstly to identify the number of points on the scale. The enquirer will, perhaps, be surprised to learn that some of the more elaborate literal scales (especially those which go in for additional symbols) have almost as many "points" on the scale as a percentage scale.

Secondly, it is necessary to be clear as to whether the scale is an equal interval scale or not. In most cases it clearly is not; it takes little to raise a 25% script to 35% but the effort needed to get a 65% script to 75% is enormous. This may not matter at one level but if marks from different scripts are being added together then it certainly does matter.

Thirdly, we have to bear in mind the distribution expected; Rowntree[12] contrasts criterion-referenced tests, in which the majority are expected to succeed, and norm-referenced tests in which the students are distributed along the dimension in something like a normal curve. Markers have in mind that few will get firsts (though, as is now well known, more get firsts in some subjects and in some institutions than in others) and few (in the British system) will fail. Universities

which teach for honours only and offer a pass degree only to those who have done badly, but not badly enough to be allowed to leave without a degree, have few in the pass degree category. There is, then, a notional distribution of final classes and this has an effect on the marks put on scripts by the markers. Should it?

Having identified how many possible points there are on the scale the next step is to discover how many points are, in practice, used. The percentage scale used in so many institutions is far from providing the 101 points one might be led to expect. The effective scale is probably nearer 50 in most cases, that is between 25 (a bad fail) and 75 (an outstanding first). Again this does not matter whilst dealing with a single group all of whom took the same paper. It does matter once scores from separate papers are conflated, especially if the students being compared in the final rank order took a different selection of the papers on offer.

The problem is now well-recognised; some papers (for example, in mathematics, science and the other subjects with a strong quantifying element) use a much larger part of the scale than others (history, English and the arts subjects generally). Furthermore the mean (the average) mark of one spread of scores may be different from that of another. Those who are interested in this aspect of marking can look up such terms as mean, mode, median, standard deviation and so forth, in any of the standard texts listed under "Further Reading" at the end of this booklet.

For the moment let us note that examiners should record

the mean score in a set of marks and the spread of the actual marks given about the mean. This is of great importance if scores from different papers with different means and different ranges of scores are to be conflated. Paper A, marked by Dr. Flint, may have a mean score of 47, a highest mark of 69 and several scores in the upper 30s. Paper B, marked by Dr. Flair, has a mean of 58, a top mark of 85 and no mark lower than 45. Even if all students took both papers injustice would still be done since strength in Dr. Flair's paper would carry more weight overall than strength in Dr. Flint's paper. There are technical ways (known as scaling) of bringing such different spreads of scores together with a common mean and a common spread but they are rarely used in the normal run of university examining. They should be. Readers are again referred to Thyne and the other standard texts listed in "Further Reading".

In fact a crude form of scaling is introduced by promulgating statements of the accepted equivalents of given classes of degree. On the percentage scale a common one is as follows:-

Numerical scale	Class of degree implied
70 and over	I
60 - 69	IIi
50 - 59	IIii
40 - 49	III
35 - 39	Pass
34 and below	Fail

This works reasonably well except for marks above 70 and below 35. If one paper, or subject, uses marks as high

as 90 whereas another never goes beyond 75, the problem remains. The same problem also occurs with "symbolic" marks of 25 or 20 for bad fails.

We shall return to this point in the next section. For the moment let us note the marks 34, 39, 49, 59 and 69 in the above list. These have become known as "terminal" marks in that they are regarded as being within the margin of error which examiners accept and which are often taken to indicate a willingness to raise the mark to the next class if the second marker or the external examiner agrees. By extension of this custom the marks 33, 38, 48 etc. are taken as indicating a firm recommendation of the class and indicates that a "lift" is not being suggested. It is even known for some Boards having decided that a mark of, say, 39 should not be raised then deliberately to lower it to the "formal" mark of 38 to indicate the firm decision taken. Perhaps this demonstrates as well as anything could the power attaching to the symbols which express our judgments.

5. COMBINING MARKS

The last section illustrated the hazards which have to be avoided when combining scores derived from different origins. The rule is quite simple: we should add like to like. We should not add apples to pears, old pence to new pence, Centigrade to Fahrenheit, or metres to yards. Yet assessment schemes do involve a large amount of combining; the questions in an essay-type paper, though marked by different hands and possibly on differing scales, have their scores conflated; course work marks are added to final examination marks; a project mark is

also thrown in and this produces a score for a unit of a degree. The rubric given to the student may say "Your work in this unit will be assessed by course work, a project and one formal examination paper of three hours, each element being equally weighted." Alas, the effective weighting may not be equal if the means and the standard deviations of the sets of marks for each element are not roughly the same.

At this stage we note briefly the question of rounding; it should be clearly stated at what stage rounding up of half marks is to occur so that practice is consistent. In the case quoted above it might be wise to have a single rounding before arriving at the final score for the unit.

If the scores, or mark, for the different elements in a unit of assessment come from subjects (or from markers) where different assumptions are being made about standards in relation to points on the marking scales used (as may happen in some multi-disciplinary courses) then it would clearly be wrong to regard "A" for one constituent element (in, for example, the scheme of assessment for the unit described above) as the equivalent of an "A" (or 70% or "Excellent") for every other element. Valuable information is often lost when scores are conflated and it is often wiser to inspect an array of scores rather than to add the scores for each element (or their supposed numerical equivalents) and divide by the number of elements in order to arrive at a single comprehensive mark for the unit. When we average in this way we are in danger of treating as like what is unlike. This problem is of increasing importance as we

move on, later in this part of the booklet, to combining scores from the separate units in a degree scheme in order to arrive at the final class of degree to be awarded.

6. COMPENSATION

The practice of adding the separate scores for, say, four questions on a paper and dividing by four assumes the principle of compensation. Giving each unit in a full degree scheme a score and then averaging the total equally assumes the principle of compensation. Extreme cases show why some guidelines are needed; what does one do with four brilliant pure mathematics papers and four marginal passes in applied mathematics? (Such a case has occurred. The ancient university concerned gave a first; perhaps the tale demonstrates that at the final stage only those who know the student are qualified to judge. The power, exceptionally, to ignore the rules should be written in!)

What is to be done with a unit assessed 50% on course work (score 70%) and 50% examination (score 29%)? The average (after rounding) comes to 50% - a lower second class degree. The simple "average the scores" rule may not always be advisable. Some systems require a stated minimum in each of the elements averaged; this can, for example, be stated in the form "The class awarded may not be more than one class above the lowest class achieved in any of the constituents". A pass in each of the constituents can also be called for.

Compensation can also come into play when a scheme has

several units of final assessment; a fail mark may be permitted provided there is compensation elsewhere. Sometimes this is defined as "double the deficiency" in terms of the additional marks needed. Another device is to permit a failed unit to be compensated provided the overall average is at a stated point higher than that needed by the student who passes all units.

Although our discussion has necessarily included compensation at the point where units of assessment are being combined to arrive at the final classification for degree awards we mean to pause here, where the scores for the elements in a unit of assessment are being combined, in order to look at the particular problems which arise in the marking of course work.

7. MARKING COURSE WORK

Increasingly schemes of assessment in higher education contain an element - sometimes substantial - of course work (sometimes called "continuous") assessment. This can be written material or contributions to seminars and other types of performance. (See Part IV). For the examiner there is a problem, referred to above, that this work will be done during the course - sometimes quite early in the course - and will then be brought alongside work produced as part of the final assessment. The standards to be applied here have therefore to be clarified. Here let it be stressed again that the work must be marked bearing in mind the conditions under which it was produced. For example, an accurate reference produced in an examination script is impressive (but not unduly so, let it be added, for the gift which some

candidates have for this may not be important in terms of the course objectives) but an exact reference in a piece of course work should be expected.

It is common experience that for a given student course work marks will be higher than terminal assessment marks, as has been noted. Many reasons could be adduced ranging from the unwillingness of tutors to face their students with a poor mark to the possibility that students get help from more able students in preparing their work, and including the fact that when we accept the "best 4 out of 6" we automatically skew the marks. Here we focus on one only of the explanations offered: it is possible that tutors carry over into course work assessment the standards they are applying in the marking of examination scripts.

In a certain sense this is right; sophisticated analysis of a critical problem can be carried out in the study as well as in the examination room. Nevertheless the possibility has to be mentioned so that we can guard against being over-influenced by the well-documented, well-illustrated case.

Who is to say whether the course work mark (assuming that there has been no unlawful assistance) is "wrong" and the examination mark "right"? Again, we need to be clear as to how we are using course work; is it to deal with aspects better dealt with in the study or the laboratory and therefore not appearing in the examination? Or is it to examine in a different way material already being assessed by formal examination?

Where course work is based on the same taught material as examinations, and where it is thought to overlap with examinations in so far as examinable skills and qualities are concerned, then it is frequently held that examination performance is what counts and course work may be given a low weighting. It may have a fixed weight or a variable weight depending on whether or not all students are to benefit from course work in proportion to its quantity and quality. It is more likely to have a fixed weight (and to have a higher weight) when it does not relate to precisely the same field of study as examinations do, or where a department is confident that course work assessment is measuring quite distinct skills or abilities. This is most clearly seen where the work is practical - as it may be in the creative arts or in the sciences.

When questions such as these are posed and answered we are a good deal of the way towards deciding how to control our marking of course work. It is certain that if we consistently give a student higher course work marks than those which will be achieved in overall assessment we have misled that student. Let us not forget that one of the purposes of marking is to provide feedback to the student, that is accurate feedback which will enable him to try to adjust his future performance.

8. MARKING OPEN BOOK AND PRIOR NOTICE EXAMINATIONS

Something of the same problem arises when marking these newer forms of examination - open book and prior notice examinations - as we face when marking course work. Again the rules are clear:-

* define the conditions carefully,
* mark in the light of those prevailing conditions.

Precision in definition is essential because there are different kinds of open book examination and the technique overlaps in some of its aspects with prior notice examinations. The simplest form of open book examination is the traditional examination with pre-determined texts ("clean copies") such as the Bible or a set book available to each candidate. Obviously the answers to questions in this type of examination will be marked with the expectation that the text will be properly and adequately referred to and that exact quotations will be used.

Prior notice papers are of two kinds; there are those given out at a stated time before the examination - the student will be expected to prepare for answers which will be written under examination conditions. (We have discussed these types of examination fully in Part IV.) The other variant is one in which the actual paper is taken away and the answers are submitted at the stated time.

This latter form of prior notice examination is no different from ordinary course work apart from the strictness of the time limits imposed for its completion. It therefore presents us with no new problems when we come to marking. The "written in the examination room" prior notice paper does have peculiarities to which we must give some thought. Some examiners take the view that a mark can properly be given for such a paper only after a viva to establish that the ideas manifested in it are those which the student now possesses and properly

understands. It is difficult to generalise; we would like simply to reiterate the need to ensure that the criteria which inform the judgments reflected in the marking are those appropriate to work carried out in conditions which are greatly different from those of the formal "unseen" examination.

9. DECLARING A CLASS OR A MARK FOR A UNIT

We have dealt with the problems of standards, of marking scales, of marking course work, and of combining the scores for the different elements in a unit of assessment. The marks (numbers, letters and symbols) for each individual item of assessed work are combined, perhaps step by step, to give increasingly more comprehensive marks. We have noted the desirability of holding on to as much as possible of the information used to arrive at combined scores, at every level, for as long as is practical in order to facilitate that wider discussion which border-line cases deserve when it comes to the classification of overall performance. For each course unit, or unit of assessment, the end result of this process is a declared class for each unit - or a declared mark which has class significance.

Examiners now come together in a Board of Examiners to consider the array of these classes (or marks) which contribute to the final degree of each student. Ideally the examiners from different areas (in, for example, a multi-disciplinary degree) will not have known until now what has been happening elsewhere. This is to guard against "halo" effects.

The task of the Board of Examiners is at its simplest when they are considering the unit marks for a pass degree; they have only to make provisional awards of "Pass" or "Fail", possibly with recommendations about the right to re-sit. Their task is more complex when degrees are to be classified.

10. CLASSIFICATION FOR FINAL AWARD

All the problems referred to in this part of the booklet occur in their most important form in the act of grading students for the purpose of declaring their final award. The sociologists speak of "labelling" and this is a good description of what the examiner does when he declares the final class of degree. Job opportunities, access to research, salary differentials are all affected by the class of degree and there is every justification for teachers in higher education giving this aspect of their profession the full attention it deserves.

There are, of course, different philosophies (if we may so describe them) of how to arrive at a declaration that an individual is to be classified thus or thus. The "credit" system sees the programme as so many units, at different levels, each examined as it is completed and the degree is then awarded on the basis of having acquired the requisite number of credits.

The more traditional (for the United Kingdom) system is to keep intermediate assessment to a minimum once entry to the substantive course has been given and to concentrate upon making an overall judgment at the end. The first system describes how the student has fared <u>en route</u>;

the second describes him on arrival.

The second system, which reviews the student's total performance at the end of the course, is the more complex in its operation. Typically the examiners will have before them a range of units, as has been noted, each with a degree class or a mark with class significance attached to it. Some way must now be found of combining these units.

The Board considers each array of unit marks or classes. They may, for example, have before them, in a 9 unit degree scheme, an array of four units in the IIi class, three in the IIii class, one III and a fail. The college or university will have rules about how to proceed and it is valuable for those who are interested in different types of rules to consider their effect on the class to be awarded. We talk at this stage of the "provisional class of degree to be awarded" because these are sighting shots made prior to aiming deliberately at a given class of degree. There are two principal ways of proceeding: the "averaging" way and the "grouping the classes" way.

Averaging to get the class

The instruction "take the average of all the units and read off the class of degree" is not as straightforward as it may seem to the beginner. If units have been given marks rather than classes one can average using the original "raw" marks; this will be a reasonable way of proceeding provided the spreads of scores in the constituent units are comparable and provided that the examiners are content to assume an equal interval scale

has been used in every case. But we must not forget about the symbolic nature of these marks and their class significance. It seems better therefore to proceed with class marks than with numerical unit scores.

If the mark for each unit is, then, expressed in classes of degree the examiners will receive an array looking something like this:-

I, IIi, IIi, IIi, IIii, III, III, Fail.

What is the average? For averaging purposes it is possible to use an artificial scale like this:-

I	IIi	IIii	III	Pass	Fail
5	4	3	2	1	0

Such a system can be fine-tuned by deciding to sub-divide each class into good, average and poor. The artificial scale would then be as follows:-

I	IIi	IIii	III	Pass	Fail
17	14	11	8	5	2
16	13	10	7	4	1
15	12	9	6	3	0

This would obviously produce a more sensitive average but again equal intervals are assumed. This scale can be "stretched" by giving greater weight to I class marks and by penalising Fail marks more heavily, thus moving away from equal intervals at the extremes. The scale would now look like this:-

I	IIi	IIii	III	Pass	Fail
22	16	13	10	7	4
20	15	12	9	6	2
18	14	11	8	5	0

The ingenious can devise even more elaborate ways of

averaging unit classes while at the same time incorporating judgments of relative value about units of different quality.

Choose your numerical scale, then, having decided whether or not to give extra weight to firsts and fails, find the average. What class of degree now? Let us return to the array

I, IIi, IIi, IIii, III, III, Fail.

You can see at once why it is sometimes permissible to discard a single Fail since it pulls down the average - too much? This rubric may have to be counteracted by another rubric, aimed at the calculating student who deliberately "throws" a unit. It must be possible for the Board to bring the failed unit back into the reckoning either when the fail is really bad or when the result in the other units produces a provisional award just above a borderline.

What else can be done? The use of a simple numerical scale equivalent to the unit classes (with a stretch at either end) for calculating a final average mark allows the examiners to construct a further table of final class equivalents for the numerical averages. Thus if the scale used was this:-

I	IIi	IIii	III	Pass	Fail
7	5	4	3	2	0

in an eight unit scheme the final class divisions might be:-

15 and below	Fail
16 - 23	Pass
24 - 31	III
32 - 39	IIii
40 - 55	IIi
?	I

But where do we put the line for a first class degree? At 48 perhaps? The problem here is that the first is at the limit of the scale and, notwithstanding the weighting built into the scale used in this case, it is difficult to avoid regression.

The scheme just outlined takes the line at which a given class begins as being the figure for the class (or the lowest figure for that class if, in a more sophisticated system, there is fine-grading within classes) multiplied by the number of units in the degree scheme. At this point anyone involved in drawing up such a scheme should work out a large number of real cases together with a large number of extreme hypothetical cases to see what actually happens.

Where simple averaging does not produce "sensible" results then the weightings (for example at the extremes), the dividing lines (between classes), and the rubrics (for example discarding one unit) need to be re-examined. If it is to work any scheme like this must operate within a set of guidelines which will encompass all but the most exceptional arrays of scores. The lines between classes need not, for instance, be set on the figure produced by multiplying the minimum by the number of units. They can be set mid-way between the points if this gives more sensible results.

It is also possible (see the section on compensation above) to vary the average to be obtained before a given classification is allowed depending on the make-up of the array of units in question. If, for example, one unit has been failed then the average needed for admission

to a class has to be higher. This is the kind of manipulation which gives the averaging procedure a bad name but it is surely obvious that the arithmetic, the rules and the procedures must never be allowed to displace the judgments being combined nor the overall judgment they are designed to facilitate.

Grouping the unit classes to get the final class

Systems which decide the class of degree by grouping unit classes must also establish a set of guidelines capable of coping with the normal range of actual cases. Commonly it will be decided that to obtain a given class the majority of the array of units must be in that class (or better) and the average of the remainder (or sometimes the remainder minus the least good unit) should be not more than one class below.

(If the reader finds it confusing that the word "average" appears in the first paragraph about the procedure we have described as the alternative to averaging he should remind himself that the distinction is a practical one and not necessarily a logical one.)

The "average" used here implies that in some cases quantification will have to take place; broad judgments may be made on the grouping of classes, finer comparisons are often easier to handle using numbers. In general the procedure for "averaging" remaining classes, once the initial grouping of classes has been considered, follows that in the preceding section. A simple numerical scale of equivalents to classes is used. This is obviously better than returning to the raw unit scores some of which may have symbolic meanings.

Grouping systems also allow for compensation by permitting a high unit (say a I when the degree being aimed at is a IIi) to pull up a lower unit (say a IIii) in order to make up the necessary majority of units within the class. There can also be additional rubrics about getting sufficient units of the final class in the major area of study. For example, a student majoring in Politics may also take courses in History and Economics. Should a provision "5 marks within a class" be tightened up to "5 marks within a class, 3 of which must be in Politics"?

At the lower end of the class list there can be rules about failures (we have touched on this already); a certain number of failed units will preclude the award of a degree; another number of failed units precludes the award of honours.

There is no end to the ingenuity which can be displayed in procedures for calculating degree classes. Some institutions use both principal ways; they calculate the average and then try "grouping by classes". Where the methods produce the same answer the provisional class of degree is awarded; where different answers emerge then the matter is debated at the Board of Examiners.

A provisional class of degree

All this figuring produces nothing more than a provisional class of degree. The Board of Examiners decide upon the exact recommendation to be made to the authorised degree-awarding body. We can be sure that there will never be a set of rules capable of providing automatic answers to the many cases presented. Rules are the guidelines for the Board of Examiners, they are to ensure that there is

a measure of consistency in the treatment of each individual case.

It is important that Boards of Examiners should have discretionary powers. Scales and averages are useful, particularly when dealing with large numbers of cases, but the figures they produce are neither magical nor sacrosanct. A Board may deliberately ignore rules provided that they draw the awarding body's attention to those recommendations which are not in line with the approved scheme; justice must be done to the occasional case which is so bizarre that it cannot fairly be treated within the scheme. Decisions in such cases will always be moderated by the judgment of an external examiner. We deal with the role of the external examiner more fully in the next section.

Meanwhile we end this section by inviting readers to declare, using any scheme they prefer, the class of degree they would award for the following patterns produced by an eight unit scheme.

Candidate	I	IIi	IIii	III	Pass	Fail
A	1	3	-	2	1	1
B	-	1	4	2	-	1
C	-	-	2	3	-	3
D	-	8	-	-	-	-

Then try your hand at writing a set of regulations designed to produce the same results when the above cases are scrutinised by those who do not share the (undeclared?) value judgments lying behind the decisions you have made.

11. THE EXTERNAL EXAMINER

A unique feature of British higher education is that all final assessment takes place with the full participation of an examiner, usually of senior status and wide experience, from another institution of standing. External examiners let the internal examiners know how they wish to proceed; some will mark "blind" a whole set of scripts and then consider the marks, given in a sealed envelope, provided by the internal examiners. More often they will sample the scripts, having been provided with a rank order, and will concentrate on borderline cases. When course work is offered it is important that the course work of all borderline candidates is available for scrutiny by the external examiner.

Internal examiners are advised to use their external examiner, drawing to his or her attention the particular problems that arise, of course, but it is <u>essential</u> also fully to use the external examiner in the normal course of examining to moderate the assessments of internal examiners. We have mentioned, more than once, the need for double-marking; the new examiner should be warned that a wide spread of marks between examiners is likely when essay questions are being marked.

Whenever doubts cannot easily be resolved in respect of particular cases the external examiner may be able to hold a <u>viva voce</u> examination, but only if this is arranged sufficiently far in advance for the candidates to be warned that they may be so examined.

Lastly, it should be noted that one of the more important

functions of the external examiner is to provide the internal examiners with comments on the work produced by the students, advice on the course being assessed and suggestions about marking and assessment. Here, too, the internal examiners may freely seek the advice of their external examiner.

12. BIAS

We began "Marking and Grading" with a section on symbols, stressing that the symbols (which have dominated recent pages) are second in importance to the judgments to which they relate. Judgments of value can be influenced by bias, conscious or unconscious, in the marker.

Rowntree,[12] quoting Husbands,[50] contrasts this "What a pedestrian attempt! Forty-five I suppose. But at least he doesn't drop a lot of leftist slogans. Okay, fifty-two" with the Marxist sociologist reading a pro-functionalist effort "Quite well argued. Sixty-two maybe. But he refuses to get to grips with the real issues. Fifty-seven".

Clearly students and examiners have points of view, orientations towards their subject and frames of reference, and these inevitably affect their reaction to what they read. We must not ignore this phenomenon.

It is indefensible if the student's mark is affected by the chance of who marks his paper. We can begin by recognising that the possibility of bias exists and training ourselves to off-set for it. We can next double mark all examination scripts, preferably blind.

In the discussions which ensue the rationale behind particular marks can be made explicit. External examiners can then be advised of our particular orientations so that a further check can be made for any unfairness which may have crept in. Course work too (especially dissertations) can and perhaps should be double-marked. In arranging for double-marking colleagues should try to arrange a partner of a different orientation; the fact that the partner may not be expert in the field is not as important as the fact of having a different point of view. A course devised, taught and assessed by one person is a closed world; we all know the dangers inherent in this.

13. PROFILES

The traditional end-product of assessment is a degree. Degrees do not, however, specify a student's strengths and weaknesses. A single figure - the degree class - tells all; many different patterns of achievement are subsumed in a single measure of quality. And when a student is told privately how well he has done he will be given little more information than an indication of the class he obtained in each of the scheduled units. Some feel that more should be told; a profile of performance is sometimes advocated as an alternative or a supplement to the single mark of the classified degree.(51,52,53)

Various kinds of profile are possible and they can be used in different ways. Here are some examples. They are not mutually exclusive; any one could be used with all or any of the others.

- A certificate of courses taken - useful for those who do not complete a degree as well as a supplement to a degree award.
- A certificate of courses taken together with a statement of the overall standard achieved in each course - including both course work and examination classifications or even the actual marks awarded.
- An aptitudes and abilities profile describing the qualities or abilities a student has manifested both during his courses and in his formally assessed work - this would be a qualitative and personal profile.
- An abilities profile consisting of a student's ratings for a number of abilities or qualities listed for this purpose by major departments.

Should such certificates and profiles be made public? With the exception of the simplest certificate it seems unlikely that many colleges or universities would agree that they should be. Should they, then, be given privately as a matter of routine to every student, along with formal notification of his degree award, to use as he wishes? These questions are well worth discussion.

A profile may certainly be of internal use - perhaps as an aid to assessment. It is difficult to measure complex abilities, especially when they cannot be measured in isolation from the content of a question or assignment, and some qualities cannot be measured at all. Perhaps a record card could be kept on which desirable qualities or abilities were listed and for which each member of staff could enter ratings, on an agreed scale,

for all the course work he examines. Something like this is done by many teacher training establishments in connection with the difficult task of assessing teaching practice. The record card could be admitted alongside marks for both course work and examinations for consideration by a Board of Examiners, especially when dealing with borderline cases. Alternatively such a record might be kept in the department to be consulted when, for example, a prospective employer applied to them for a reference for any student. Profiles are particularly useful when a candidate who has taken joint or combined honours applies for admission to a research degree in one of the subjects of his first degree.

A fuller discussion of the practical value of profiles and certificates - on the one hand to individual students and on the other hand to prospective employers or other interested bodies - will be of interest to those who feel that alternatives to or modifications of traditional degree classes should be explored. (Klug[51] gives a very full account of the merits of profiles though it is coupled with an attack on the exclusive use of classified degrees. Douglass[54] discusses Learned's full academic inventory profile.)

PART VI

CONCLUSION

There is a considerable literature on assessment and the final pages of this booklet provide the interested teacher with a list of references which are both of general interest and which are capable of providing a basis for further study. The fact that the literature is growing rapidly is an indication that scholars are everywhere concerned to improve the means used to assess students.

There is scope for innovation and it is to be hoped that further experimentation will be planned, executed and reported. Those who have read thus far are invited to analyse carefully the assessment procedures used in their own institutions, noting the assumptions made, reviewing the rules employed, calculating the effect of different sets of rules and, above all, estimating the effect of the assessment procedure on the learning which a) takes place and which b) is intended to occur. Learning is at the centre of a college or university; the teaching is to provoke learning and the examinations are there to assess it.

Considerations about what shall be tested, by what methods and when, are important and must be decided honestly, fairly and openly. But none of these questions is as important as those concerned with the learning achieved by the students. That is why the methods of assessment should in all cases flow from a statement of learning objectives; if the reverse begins to happen then we know that it is time to review our assessment procedures.

FURTHER READING

1. Discussion papers, descriptive critiques and accounts of assessment in principle and in practice.

NUFFIELD GROUP for Research and Innovation in Higher Education (1975) *A Question of Degree: assorted papers on assessment.* London. Nuffield Foundation.

ROWNTREE, D. (1977) *Assessing Students: How Shall We Know Them?* London. Harper & Row. (An illuminating study with a good bibliography)

BLIGH, D.A., EBRAHIM, G.J., JAQUES, D. and WARREN PIPER, D. (1975) *Teaching Students:* (Section on "Assessment of Students" pp.43-54) Exeter. Exeter University Teaching Services.

COX, R. (1967) "Examinations and higher education: a survey of the literature" *Universities Quarterly 21:* pp.292-340. (This is a good survey particularly of studies on the reliability and validity of traditional examinations. It makes a case for using objective tests.)

CHARVAT, T. et al. (1968) *A Review of the Nature and Uses of Examinations in Medical Education.* World Health Organisation.

FLEMING, P.R. et al. (1976) *Examinations in Medicine.* Edinburgh. Churchill Livingstone.

2. Introductions to the principles of examining and the statistics of marking.

THYNE, J.M. (1974) *Principles of Examining.* London. University of London Press.

TITTLE, C.K. and MILLER, K.M. (1976) *Assessing Attainment.* London. Independent Assessment and Research Centre.

HUDSON, B. (1973) *Assessment Techniques* London. Methuen.

PIDGEON, D.A. & YATES, A. (1968). *An Introduction to Educational Measurement.* London. Routledge and Kegan Paul.

3. More advanced treatments of educational measurement.

THORNDIKE, R.L. and HAGEN, E. (1961) (3rd edn) *Measurement and Evaluation in Psychology and Education* New York. John Wiley and Sons.

VERNON, P.E. (1956) (2nd edn) *The Measurement of Abilities.* London. University of London Press.

LAUWERYS, J.A. and SCANLON, D.C. (eds.) (1969) *World Yearbook of Education 1969: Examinations* London. Evans.

REFERENCES

1. MILLER, C.M.L. & PARLETT, M. (1974) *Up to the Mark: A Study of the Examination Game*
   London. S.R.H.E.
2. HARRIS, N.D.C. (1975) "What is assessment?"
   Assessment in Higher Education 1 (1): pp.5-12.
3. FURNEAUX, R.D. (1962) "The psychologist and the university"
   Universities Quarterly 17: pp.33-47.
4. THYNE, J.M. (1974) *Principles of Examining*
   London. University of London Press.
5. RIPPERE, V.L. (1974) "On the validity of university examinations"
   Universities Quarterly 28: pp.209-218.
6. BLOOM, B.S. (ed.)(1956) *Taxonomy of Educational Objectives: the classification of educational goals. Handbook I: Cognitive Domain.*
   London. Longmans.
7. BILLING, D.E. (1973). "Classification of educational objectives"
   in BILLING, D.E. & FURNISS, B.S. (eds.) (1973) *Aims, Methods and Assessment in Advanced Science Education*
   London. Heyden.
8. KRATHWOHL, D.R., BLOOM, B.S. & MORIA, B.B. (1964) *Taxonomy of Educational Objectives: the classification of educational goals. Handbook II: Affective Domain*
   London. Longmans.
9. EBEL, R. (1965) *Measuring Educational Achievement*
   New York. Prentice Hall.
10. BEARD, R.M., HEALEY, F.G., & HOLLOWAY, P.J. (1974) *Objectives in Higher Education* (2nd. edition)
    London. S.R.H.E.
11. Conference convened by the Committee of Vice-Chancellors and Principals and the Association of University Teachers. (Spring 1969).
12. ROWNTREE, D. (1977). *Assessing Students: How Shall We Know Them?*
    London. Harper & Row.
13. BOURDIEU, P. & PASSERON, J.C. (1970). *La Reproduction*
    Paris. Editions de Minuit.
14. N.U.S. (1969) Executive report on examinations: submitted to April conference of National Union of Students, 1969.
15. LOVELL, K. (1968) "Examinations and Marking"
    In LAYTON, D. (ed.) *University Teaching in Transition*
    Edinburgh. Oliver & Boyd.
16. SIMPSON, M.A. (1972) *Medical Education: a critical approach*
    London. Butterworth.
17. MALLESON, N.(1959) "Evading failure by university students".
    The Listener, March 5th.
18. ABRAHAMSON, S. (1966) "A study of the objectivity of the essay examination"
    Journal of Medical Education 39: pp.65-68.
19. KNOX, J.D.E. (1975) *The Modified Essay Question (Booklet No.5)*
    Dundee. Association for the Study of Medical Education.

20. MOWBRAY, R.M. & DAVIES, B.M. (1967) "Short note and essay examinations compared"
British Journal of Medical Education 1: pp.356-358.
21. KRARUP, N. et al. (1974) "Open book tests in a university course"
Higher Education 3: pp.157-164.
22. JEHU, D. et al. (1970) "The use of notes in examinations"
British Journal of Educational Psychology 40: pp.335-337.
23. *Assessment of Students* (1975) Reprint of 1969 Report of the working party on assessment of students.
University of Sheffield. Academic Development Committee.
24. BEARD, R.M. (1972) *Teaching and Learning in Higher Education* (2nd edition)
Harmondsworth. Penguin.
25. FELDHUSEN, J. (1961) "An evaluation of college students' reactions to open book examinations"
Educational and Psychological Measurement 21: pp.637-646.
26. FLOOD PAGE, C. (1967) "Worrying about examinations"
Cambridge Institute of Education Bulletin Vol. 3, No.6: pp.2-7.
27. BLIGH, D.A., EBRAHIM, G.J., JAQUES, D. & WARREN PIPER, D. (1975)
Teaching Students
Exeter. Exeter University Teaching Services.
28. FLEMING, P.R. et al. (1976) *Examinations in Medicine*
Edinburgh. Churchill Livingstone.
29. VERNON, P.E. (1963) *The Measurement of Abilities* (4th impression)
London. University of London Press.
30. LENNOX, B. (1974) *Hints on the setting and evaluation of Multiple Choice Questions of the One-from-Five type (Booklet No.3)*
Dundee. Association for the Study of Medical Education.
31. JACKSON, D. & JAQUES, D. (eds.) (1976) *Improving Teaching in Higher Education.*
London. U.T.M.U.
32. INKSTER, I. (1971) "Some criticisms of a coursemark system"
Higher Education Review 3 (3): pp.45-48.
33. CLEMENTS, R.V. (1967) "Degree class assessment for arts and social science"
Times Educational Supplement No. 2726: p.283
34. BROCKBANK, P. (1968) "Examining exams"
Times Literary Supplement No.3465: pp.781-782.
35. NUFFIELD Group for Research & Innovation in Higher Education (1975) "Reworked Essays: an alternative method of assessment to examinations" A case-study. (The Department of German Studies, University of Warwick)
In *A Question of Degree*
London. Nuffield Foundation.
36. Reply by Department of Politics to Questionnaire in survey for the Committee on Examinations, University of Lancaster (1970).
37. WHITELAND, T.W.R. (1966) "The selection of research students"
Universities Quarterly 21: pp.44-47.
38. "Guidelines on dissertations" School of Education, University of Lancaster.

39. ST. JAMES-ROBERTS, I. (1976) "Cheating in science"
New Scientist, 72: (1028): pp.466-469.
40. NUFFIELD Group for Research & Innovation in Higher Education (1975)
Discussion Paper I
In *A Question of Degree*
London. Nuffield Foundation.
41. ADDERLEY, K. et al. (1975) *Project Methods in Higher Education*
London S.R.H.E.
42. HEYWOOD, J. (1969) "Assessment of undergraduate performance"
Background paper prepared for the Universities Conference convened by the Committee of Vice-Chancellors and Principals and the Association of University Teachers.
43. ENTWISTLE, N.J. & WILSON, J. (1977) *Degrees of Excellence: The Academic Achievement Game*
London. Hodder & Stoughton.
44. JONES, P.M. (1965) "The fallacies of the examination system"
Universities Quarterly 19: pp. 240-258.
45. NUFFIELD Group for Research & Innovation in Higher Education.
Nuffield Newsletters 1-7 (1973-1976)
46. McGUIRE, C.H. (1963) "A process approach to the construction and analysis of medical examinations"
Journal of Medical Education 38: pp.556-559.
47. BLACK, B.J. (1968) "University examinations"
Physics Education 3: pp.93-99.
48. BEARD, R.M. & POLE, K.E. (1971) "Content and purpose of biochemistry exams"
British Journal of Medical Education 5: pp.13-21.
49. CHAPMAN, L.R. & BUTLER, P. (1971) "July thoughts"
Mathematics Gazette 55: pp.22-28.
50. HUSBANDS, C.T. (1976) "Ideological bias in the marking of examinations: a method of testing for its presence and its implications"
Research in Education No. 15: pp.17-38.
(As quoted by Rowntree, D. (1970) *Assessing Students: How Shall We Know Them?* London. Harper & Row.)
51. KLUG, B. (1974) *Pro Profiles*
London. N.U.S. Publications.
52. KLUG, B. (1977) *The Grading Game*
London. N.U.S. Publications.
53. HINTON, D. (1975) "The evolution of the Birmingham Profile; a background paper (1973)
in NUFFIELD Group for Research & Innovation in Higher Education
A Question of Degree
London. Nuffield Foundation.
54. DOUGLASS, P. (1960) *Teaching for Self Education; as a life goal*
New York. Harper Brothers.

TEACHING IN HIGHER EDUCATION SERIES

1. Lecturecraft
2. Why Lecture?
3. Working Together
4. Participatory Discussion
5. A Kind of Learning
6. Patterns and Procedures
7. Inside Assessment
8. Essays
9. The Use of Objective Tests

School of Education, University of Lancaster